

What is Claimed Is

- [c1] A method of converting a document in a page-image format into a form suitable for an arbitrarily sized display, comprising:
- deconstructing a document in a page image format;
 - synthesizing the deconstructed document into an intermediate data structure; and
 - distilling the intermediate data structure for redisplay in a format usable for an arbitrarily sized display.
- [c2] The method of claim 1, wherein deconstructing a document in a page image format includes:
- identifying text image areas and non-text image areas of the document;
 - locating and isolating text image areas and non-text image areas;
 - processing the isolated text image areas and non-text image areas into text line regions and layout properties;
 - processing located text line regions into segmented image elements; and
 - locating and labeling segmented image elements.
- [c3] The method of claim 2, wherein deconstructing a document in a page image into the set of segmented image elements includes at least one of physical segmentation of data and logical segmentation of data.
- [c4] The method of claim 2, wherein the set of segmented image elements comprises at least one of blocks, lines, words, characters of text, groups of characters, and groups of non-text characters.
- [c5] The method of claim 1, wherein synthesizing includes converting non-text image areas, layout properties and segmented image areas into the intermediate data structure.
- [c6] The method of claim 2, wherein synthesizing the set of segmented image elements into an intermediate data structure includes integrating at least one of bitmapped images in an intelligible display layout and links to non-textual elements.
- [c7] The method of claim 6, wherein the bitmapped images are images of words in

reading order.

- [c8] The method of claim 1, wherein the intermediate data structure is stored in a storage device.
- [c9] The method of claim 1, wherein distilling the intermediate data structure for redisplay in a format usable for an arbitrarily sized display, includes redisplaying the document in human readable format
- [c10] The method of claim 1, wherein distilling the intermediate data structure for redisplay in a format usable for an arbitrarily sized display, includes redisplaying the document in at least one of an electronic book format, Internet browsable format and a print format.
- [c11] The method of claim 1, wherein distilling the intermediate data structure includes converting the stored intermediate data structure into a device specific display format for display.
- [c12] The method of claim 1, wherein the intermediate data structure is adaptable to at least one of display screen size, page size, resolution, contrast, color and geometry, at the time of display.
- [c13] The method of claim 1, wherein the intermediate data structure is adaptability supported by at least one of repagination of text, reflowing of text, logical links of text to associated text and non-textual content.
- [c14] A method of converting a document in a page-image format into a form suitable for an arbitrarily sized display, comprising:
analyzing page layout;
converting a sequence of page images into a sequence of document element images captured in a tagged format; and
re-converting the tagged format into at least one of an electronic book format, an Internet browsable format that can accept images and a print format.
- [c15] The method of claim 14, wherein the tagged format preserves at least one of reading order and logical page layout properties.

~

[c16] A system of converting a document in a page-image format into a form suitable for an arbitrarily sized display, comprising:

- an input/output device;
- a controller;
- a deconstructing circuit, routine or application that deconstructs a document;
- a synthesizing circuit, routine or application that synthesizes the deconstructed document into an intermediate data structure;
- a distilling circuit, routine or application that distills the intermediate data structure for redisplay in a format usable for an arbitrarily sized display;
- a memory.

[c17] The system of claim 16, wherein:

- the deconstructing circuit, routine or application deconstructs the document in a page image format into non-text image areas, layout properties, and a set of segmented image elements;
- the synthesizing circuit, routine or application synthesizes the non-text image areas, the layout properties, and the set of segmented image elements into an intermediate data structure; and
- the distilling circuit, routine or application distills the intermediate data structure for redisplay in a format usable for an arbitrarily sized display.

[c18] The system of claim 17, wherein the deconstructing circuit, routine or application deconstructs the document in a page image format into the set of segmented image elements that includes at least one of physical segmentation of data and logical segmentation of data.

[c19] The system of claim 17, wherein the intermediate data structure includes at least one of bitmapped images in an intelligible display layout and links to non-textual elements.

[c20] The system of claim 19, wherein the bitmapped images are images of words in reading order.

[c21] The system of claim 16, wherein the memory stores at least one of the

document in page image format, the deconstructed document, the intermediate data structure and the distilled document.

- [c22] The system of claim 16, wherein the distilling circuit, routine or application distills the intermediate data structure for redisplay of the document in a format usable for an arbitrarily sized display includes redisplaying the document in at least one of an electronic book format, Internet browsable format, and a print format.
- [c23] The system of claim 16, wherein the distilling circuit, routine or application converts the stored intermediate data structure into a device specific display format for display.
- [c24] The system of claim 16, wherein the intermediate data structure is adaptable to at least one of display screen size, paper size, resolution, contrast, color and geometry, at the time of display.
- [c25] The system of claim 16, wherein the intermediate data structure is adaptability supported by at least one of repagination of text, reflowing of text, logical links of text to associated text and non-textual content.
- [c26] The system of claim 16, wherein the deconstructing circuit, routine or application analyzes page layout and converts a sequence of page images into a sequence of document element images captured in a tagged format; and
the distilling circuit, routine or application converts the tagged format into at least one of an electronic book format, an Internet browsable format that can accept images and a print format.
- [c27] The system of claim 26, wherein the tagged format preserves at least one of reading order and logical page layout properties.
- [c28] The system of claim 26, wherein the deconstructing routine includes a segmentation algorithm and a background structure analyzer.